

PSEUDORABIES VIRUS GLYCOPROTEIN E GENE: SEQUENCE ANALYSIS AND RELATIONSHIP TO OTHER HOMOLOGOUS GENES

N. A. Zeenathul¹, M. L. Mohd-Azmi^{1,2}, A. R. Sheikh-Omar¹, I. Aini¹,
A. R. Bahaman¹ and A. R. Mutalib¹

¹Faculty of Veterinary Medicine, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia

²Institute of Bioscience, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia

SUMMARY

The glycoprotein E (gE) gene is a virulence-associated gene of pseudorabies virus (PrV). This paper reports the first documented gE sequence analysis of a locally derived PrV. A comparative sequence analysis with other herpesviral gE homologues was also performed to give an insight on where it stands among the pool of genes. The gE gene of TK⁻gE⁺ PrV features a typical type 1 membrane protein which starts with the initiator methionine and followed directly with a 27-amino-acids signal sequence. Basically the gene could be divided into three distinct functional domains: a 429-amino-acid ectodomain, a 26-amino-acid hydrophobic transmembrane domain and a 123-amino-acid, highly charged cytoplasmic domain. A high degree of gE gene conservation was shared between TK⁻gE⁺ PrV and other PrV strains. Only 23 to 31% homology was found when compared among the gE proteins of alphaherpesvirus from diverse animals. Despite the low overall level of identity, considerable similarity of cysteine rich regions was observed among gE genes, indicating some important sequences for the structure of these glycoproteins. Although, the gE of PrV strains have closely conserved predicted N-linked glycosylation sites, they have no counterparts in the homologous proteins of other alphaherpesviruses. Comparison of amino acid sequences with gE homologs indicated a greater diversity of sequence in the N- terminal region of the protein. It also highlighted several features of the gE protein conserved throughout the herpesvirus family which is also shared by TK⁻gE⁺ PrV. Characterisation of the local PrV gE at molecular level may facilitate the construction of recombinant or chimeric PrV as vehicles for the delivery of vaccine antigens to the host.

Keywords: Pseudorabies virus, glycoprotein E gene, sequence analysis, herpesviral gE, homologues

INTRODUCTION

Pseudorabies virus (PrV) is a neurotropic alphaherpesvirus that causes acute fatal disease in a variety of mammals and birds. The pig is the only natural host for the virus (Enquist *et al.*, 1999). The disease is characterised by a variety of clinical signs in older pigs, including encephalitis, pneumonia, increased susceptibility to other respiratory pathogens and reproductive complications. The most common mode of transmission of the virus is via direct contact with infected pigs and infected body secretion (Schoenbaum *et al.*, 1990). Although PrV has many genes that contribute to viral infectivity, glycoprotein E (gE) is one of the most studied genes because of its importance as a virulence mediator and dissemination of PrV in every animal model tested thus far (Tirabassi and Enquist, 1999). Previous findings by Card *et al.* (1992) demonstrated that gE is important for the transneuron transport of PrV in rats, while Kimman *et al.* (1992) suggested that gE is important for the transport of PrV through the porcine CNS. Meanwhile, Jacobs *et al.* (1993a) found that deleting valine and cysteine from gE reduced the size of plaques and reduced the virulence for mice to the same degree as deleting the entire gE protein. Another study by Tirabassi *et al.* (1997) reported that the N-terminal extracellular

domain of gE is sufficient to mediate gE-promoted spread in the rat central nervous system (CNS), while the C-terminal cytoplasmic domain of gE is required for gE mediated virulence. However, the mechanism by which gE accomplishes these two separable functions has not yet been determined. Therefore, more work need to be done to give clues for related functions of the gene.

In the present study, the gE gene sequence of a locally derived TK⁻gE⁺ PrV strain (Zeenathul, 1999) was molecularly characterised. Since PrV has similar genes with other herpesvirus, a comparative gE sequence analysis was also performed to determine the degree of gene conservation among gE homologues. The data set from this study may facilitate the construction of chimeras or recombinant PrV as vehicles for the delivery of vaccine antigens to the host.

MATERIALS AND METHODS

Virus

The PrV isolate (TK⁻gE⁺ PrV) considered in this study has been described previously by Zeenathul (2004). The TK defective (TK⁻) PrV which had been derived from a local virulent strain, was established at the Virology Laboratory, Faculty of Veterinary Medicine,

Universiti Putra Malaysia (Zeenathul, 1999). The virus was propagated in Vero cells and viral DNA was extracted following established methods (Zeenathul, 2004).

PCR amplification of the full length PrV gE gene

Primers used to amplify the gE gene were designed based from the published gI and gE sequences (Petrovskis *et al.*, 1986) and modification of previously established gE primers (Jacobs *et al.*, 1993b). Forward primer, PrVgef (5'-CAAGATGACGTTGGCCGAGCTTCG-3') is located upstream at nts -161 to -138 while the reverse primer R-GE11K (5'-CGGAATGCGGGAGGACCGGTTATCC-3') extends downstream towards 11K gene (nts -3 to -28 of 11K gene). PCR was carried out in a 50 μ l reaction mixture with an established thermal cycling protocol. An initial pre-denaturation of viral DNA at 99°C for 3 min was performed prior to addition of the remaining cocktail mixture. The mixture was heated at 95°C for 2 min followed by 30 cycles of 1 min at 94°C, 1 min at 65°C, and 1 min at 72°C, with a 5-min final extension at 72°C. PCR-amplified DNA fragments were run on 1% agarose gel electrophoresis and the gel was purified by QIAquick Gel Extraction Kit (Qiagen, U.S.A.). The fragment was subsequently cloned in PCR[®]-Blunt vector (Invitrogen, USA).

Sequence analysis

The cloned gE gene was sequenced in an automated DNA sequencer (ABI systems, USA). Sequencing was done in both directions using M13 forward and reverse universal primers as well as gE specific primers for walking along the gE gene as follows: (forward SPECFO1 5'-TGC GAC GCC GTG GCG GTG ACC A-3' (nts 348 to 369);

forward PrVE05 5'-ACTACGTGTACGAGCCCTGCA TC-3' (nts 829 to 851); reverse E04 5'-GTG GGC AGG CTG GTG TAC ACC GGC G-3' (nts 1422 to 1447) and reverse geR 5'-TAA GCG GGG CGG GCA TTC AAC AGG-3' (nts 1732 to 1709). SDSC Biological WorkBench 3.2 was utilised to analyse the nucleotide sequences and deduced amino acid sequences. For sequence comparison, GenBank database was used. The GenBank accession numbers of the sequences used for analysis were bracketed as follows: PrV:Ea strain [AF171937.1], PrV:Rice strain [P08354], Canine herpesvirus (CHV) [AAB67060.1], Bovine herpesvirus 1 (BHV-1) [Q08101], Equine herpesvirus 1 (EHV-1) [M36299.1], Herpes simplex virus (HSV-1) [P04488], Gallid herpesvirus 2 (GaHV-2) [CAA48619.1] and Simian varicella virus (SVV) [Q04548]. Studies on homology search (as predicted by Statistical Analysis of Protein Sequence), phylogenetic relationship (Phylip's Drawtree), multiple sequence alignment (CLUSTAL W, TMAP), FASTA search, and codon changes as well as amino acid determination were included.

RESULTS

The G+C rich PrV DNA hampered PCR amplification and led to laborious optimisation. In this study, various PCR additives were tried; however, only DMSO was helpful in amplifying a correct-sized DNA fragment from the PrV DNA. Usage of heavy predenaturation up to 99°C (Wang Haijian, Nanjing Agricultural University, Jiangsu, China, pers.comm.) had partially solved the problems. Considerable specificity was obtained when the temperature of the primer annealing step was within 65°C-70°C. Despite optimisation, multiple fragments were generated during PCR[®]. Therefore, the amplification

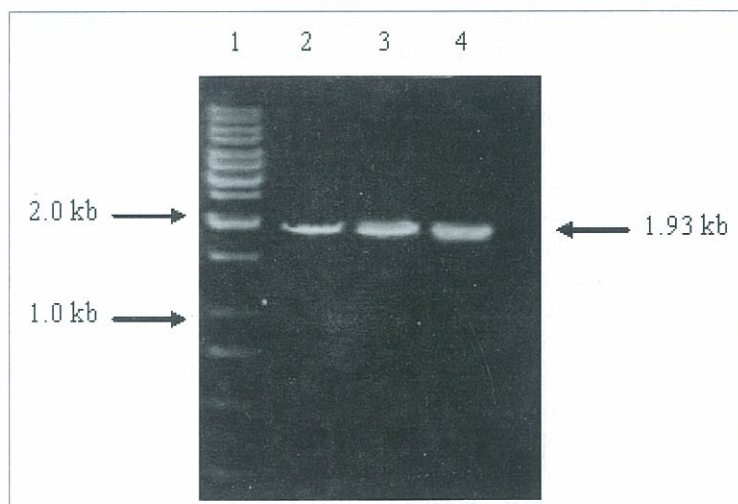


Figure 1: Purified PCR product of the TK⁻gE⁺ PrV gE gene. The PCR product of 1.93 kb containing the full-length gE sequence was amplified using PrVgef (sense) and R-GE11K (antisense) primers. Lane 1: 1kb DNA ladder (Fermentas); Lanes 2 to 4: gel purified PCR product.

product of 1928 bp was gel purified prior to cloning in pCR -Blunt plasmid for sequencing (Figure 1).

The full-length gE sequence of TK⁻gE⁺ PrV is shown in Figure 2. The sequenced 1928 nucleotides (n) constituted the terminal end of gI gene, followed by a non-coding region, open reading frame (ORF) of gE, terminal end of gE, and the beginning of the adjacent 11K gene. The ORF encodes a putative protein of 578 amino acids with a calculated molecular weight of 62.7 kDa in unmodified form. The gE gene which features a typical 1 membrane protein (Tirabassi *et al.*, 1997), starts with the initiator methionine (ATG) (Figure 2) and followed directly with a 27-amino-acids signal sequence. Basically the gene could be divided into three distinct functional domains: a 429-amino-acid ectodomain, a 26-amino-acid hydrophobic transmembrane domain and a 123-amino-acid, highly charged cytoplasmic domain (Figure 2). The length of the entire coding region of TK⁻gE⁺ PrV gE is identical to the PrV Ea strain but differs from the Rice strain (Table 1).

Multiple sequence alignment of gE homologous protein was achieved by employing the CLUSTAL W of SDSC Workbench 3.2 software allowing a Gap open penalty of 10 and a Gap extension penalty of 0.20 (Figure 3). As it is true for the entire PrV genome (Petrovskis *et al.*, 1986), the gE gene is extremely rich in guanine and cytosine. The gE genes of TK⁻gE⁺ PrV, Ea and Rice strains have G+C contents of 73%, 74 % and 74.8 %, respectively (NASTATS tool in Workbench). The codon usage in these genes favour codons with G or C in the third position. In the gE of TK⁻gE⁺ PrV, 97 % of the codons have a G or C in the third position as depicted from the Codon Usage Database (Source: GenBank Release 138). Amino acids with the G + C-rich codons are very abundant in the gE protein. The predicted gE of TK⁻gE⁺ PrV has 11 % alanine (A), 11% proline (P), 8% glycine (G), and 8% arginine (R) which is similar to that of Ea strain (Figure 3). However, the Rice strain differed in

the alanine composition (12%). Among the gE homologous, HSV-1 showed the highest percentage of alanine (16%) and proline (12%) while CHV, the least (2.3% and 4.6, respectively). The range of glycine and arginine composition among the homologues was 3.3-8% and 2.3-8%, respectively. Overall, the composition of these GC rich coding amino acids (aa) varied among the gE homologues.

Homology search revealed 98% of aa similarity within the TK⁻gE⁺ PrV and the Ea PrV strain gE gene sequences (Table 1; Figure 4). Out of ten nucleotide variations (at n position 237, 931, 1207, 1409, 1501, 1530, 1549, 1555, 1682 and 1842, respectively), six sites coded for non identical amino acids. The most divergence was found in the C-terminal region, consisting of the 5 amino acid (aa) substitutions at aa position 470 (V6A); aa 501 (V6I); aa 517 (P6S); aa 519 (T6A) and at aa 561 (T6N) except for one specific region located at aa 403 (A6P) (Figure 3). The presence of an aspartic acid (D) at aa position 128 of the gE coding region, corresponded to an addition to the aa abundance in TK⁻gE⁺ PrV and Ea strain (Table 1). When comparison was performed with PrV Rice strain, 95 % homology was identified (Table 1). However, both TK⁻gE⁺ PrV and Ea strain showed significant discrepancies with Rice gE sequence, either in terms of nucleotides or amino acids variations which are scattered throughout the reading frame without particular region preferences (Figure 3). The ranges of sequence identity depicted from homology search to other gE homologous protein are as shown in Table 1.

Ten extremely conserved cysteine clusters which aligned perfectly at aa 118 (C₁), 127 (C₂), 132 (C₃), 142 (C₄), 275 (C₅), 284 (C₆), 293 (C₇), 301 (C₈), 320 (C₉) and 332 (C₁₀), were presented within the gE PrV strains (Figure 3). A considerable degree of cysteine conservation was also observed when compared to the rest of gE counterparts (Figure 5). Interestingly, the above mentioned cysteine clusters that stretched along aa 275 to aa 332 seemed to

Table 1: Comparison of gE glycoprotein of TK⁻gE⁺ PrV with other homologous proteins of the *alphaherpesvirinae*

Virus	Acession (GI:Genbank Identification)	Number of amino acids	MW KDa	% identity of amino acids #
Local TK ⁻ gE ⁺ PrV	-	578	62.7	100
PrV: Ea strain	AF171937.1 (GI:5764548)	578	62.6	98
PrV:Rice strain	P08354 (GI:138245)	577	62.3	95
Canine herpesvirus (CHV)	AAB67060.1 (GI:2337934)	522	59.9	25
Bovine herpesvirus 1 (BHV-1)	Q08101 (GI:1174953)	575	61.2	31
Equine herpesvirus 1 (EHV-1)	M36299.1 (GI:330787)	550	61.2	28
Herpes simplex virus (HSV-1)	P04488 (GI:138240)	550	59.1	28
Gallid herpesvirus 2 (GaHV-2)	CAA48619.1 (GI:406791)	498	55.1	23
Simian varicella virus (SVV)	Q04548 (GI:549306)	604	67.6	25

1	M R P F L L R A A Q L L A L L A L A L S	60
	atgcgccgcccccttctgctgcgcgccgcgagctcctggcgctgctggccctggcgctctcc	
61	T E A P S L S A E T T P G P V T E V P S	120
	accgaggccccgagcctctccgcccagacgacccccggcccccgtaaccgaggtcccgagt	
121	P S A E V W D D L S T E A G D D D L N G	180
	ccctcgccgaggtctgtggagacacctctccaccgagggccggcgacgatgacctcaacggc	
181	D L D G D D R R A G F G S A L A S L R E	240
	gacctcgacggcgacgaccgcccgcggggcttcggctcgccctcgccctccctgagagag	
241	A P P A H L V <u>N V S</u> E G A <u>N F T</u> L D A R	300
	gcgcccccgcccatctgtgacgtgtccgagggcgccaacttcacctcgacgcgcg	
301	G D G A V L A A G I W T F L P V R G C D A	360
	ggcgacggcgccgtgctggcgccggatctggacgttctcgccgctccgcggtcgacgccc	
361	V S V T T V C F E T A C H P D L V L G R	420
	gtgtcggtgaccacgggtgtgtcttcgagaccgctgccaccggaactggtgtggcgcg	
421	A C V P E A P E M G I G D Y L P P E V P	480
	gcctcgctccccgagggccccggagatgggcatcggcgactacctgcccccgaggtgccg	
481	R L R R E P P I V T P E R W S P H L S V	540
	cggctccggcgcgagccgcccacgtcaccccgagcggtggctcgccgcacctgagcgctc	
541	L R A T P <u>N D T</u> G L Y T L H D A S G P R	600
	ctgccccgacgccccaacgacacgggcctctacacgctgcacgacgctcgggcgccgg	
601	A V F F V A V G D R P P A P A D P V G P	660
	gccgtgttcttctgtggcggtggcgacggcgcccgccgccccggcgaccccggtgggcccc	
661	A R H E P R F H A L G F H S Q L F S P G	720
	gcgccccacgagccccgcttcacgctcggttccaactcgcaactcttctcgccccgg	
721	D T F D L M P R V V S D M G D S R E <u>N F</u>	780
	gacacgttcgacctgatgccgcgctggtctcgacatggcgactcgcgcgagaacttt	
781	<u>T</u> A T L D W Y Y A R A P P R C L L Y Y V	840
	accgccacgctggactggtactacgcgcgcgcccccggtgctgtgtactacgtg	
841	Y E P C I Y H P R A P E C L R P V D P A	900
	tacgacctgcacatcacccccgcgcgcgcgagtgctcgccccggtggaccgcg	
901	C S F C T S P A R A R L V A R R A Y A S C	960
	tgcagcttcacctcgccggcgcgcgcggttggtggcgcgcgcggtacgcctcgtgc	
961	S P L L G D R W L T A C P F D A F G E E	1020
	agcccgctgctcggggaccggtgggtgacgcctgcccttcgacgccttcggcgagag	
1021	V H T <u>N A T</u> A D E S G L Y V L V M T H N	1080
	gtgcacagaaagccacggcgagtgctgggctgtacgtgtcgtgatgacccaaac	
1081	G H V A T W C D Y T L V A T A A E Y V T V	1140
	ggccacgtcgccacctgggactacacgctcgcgccaccgccccgagtagctacggtc	
1141	I K E L T A P A R A P G T P W G P G G G	1200
	atcaaggagctgacggccccggccccggccccggcaccccgtagggccccggcgcg	
1201	D D A I Y V D G V T T P A P P A R P W N	1260
	gacgacgcatctacgtggacggcgctcagacgcggcgccccggcgccccgtggaac	
1261	P Y G R T T P G R L F V L A L G S F V M	1320
	ctgtacggccggacgacgcggcggtgtttgtgctggcgctgggctccttcgtgatg	
1321	T C V V G G A V W L C V L C S R R R A A	1380
	acgtgctgctggggggcgctctggtctgctgctgctcccgcgccggcgcc	
1381	S R P F R V P T R V R T H M L S P V Y T	1440
	tcgcggcgcttcgggtgcccagcggggtgcggagcacatgctctctccggtgtacacc	
1441	S L P T H E D Y Y D G D D D G E A A G V	1500
	agcctgccacgcacgaggactactacgacggcgacgacgacgagggcggtc	
1501	V R R R P A S P G G D S G Y E G P Y T S	1560
	gtccgccccggcccccgctccccggcggggacagcggtacgaggggccccgtacacgagc	
1561	L D P E D E F S S D E D D G L Y V R P E	1620
	ctggacccccgaggacgagttcagcagcgacgaggacgacgggctgtacgtgcgccccgag	
1621	E A P R S G F D V W F R D P E K P E V T	1680
	gaggccccgctccggtccgacgtctggtccgcgacccggagaaaccggaagtgcg	
1681	T G P N Y G V T A N R L L M S R P A *	1737
	actggaccaactatggcgtagaccgccaaccgctgttgatgtcccgccccgcttaa	

Figure 2: Nucleotide sequence and deduced amino acid sequence of the gE gene of (TK⁻gE⁺ PrV). The sequence begins with an initiator methionine (at position 1). The stop codon is indicated as *. The asparagines which may be the potential sites of N-linked glycosylation are underlined.

A horizontal number line starting at 0 and ending at 50. Major tick marks are labeled at 10, 20, 30, 40, and 50. Minor tick marks are present every 1 unit. Asterisks (*) are placed above the number line at positions 5, 15, 25, 35, and 45.

MRMTVVVKHVMLTLICGTLSWGVOINTLAYASAIKSEDEGFDMDGCVYGD

-----MRPFLLRAAQLLALLALALSTEAPSLSAETTPG
-----MRPFLLRAAQLLALLALALSTEAPSLSAETTPG
-----MRPFLLRAAQLLALLALALSTEAPSLSAETTPG
-----MFLFFLIASITLCSVRFGSTIRN
-----MQPTAPRRRRLPLLQLLFLGLMAEAPATETPGSA
-----MELLAASRACIFFGLVTVLDAGVQQ
-----MDRGAVVGFLGVCVVSCLAGTPEKT
-----MRACRLSILILLQTISAMQH
IODYINAA¹Y²TH³ER⁴PF⁵IOD⁶SK⁷HK⁸ME⁹TY¹⁰SL¹¹HT¹²PLE¹³TD¹⁴LS¹⁵Q¹⁶IS¹⁷IR¹⁸EG¹⁹NY²⁰

PVTEVPSPSAEVWDDLSTEAGD----DDLNGDLGDGDRRAGFGSALASL
 PVTEVPSPSAEVWDDLSTEAGD----DDLNGDLGDGDRRAGFGSALASL
 PVTEVPSPSAEVWD-LSTEAGD----DDLGDGLNGDGRRAGFGSALASL
 --VIVSEKSGFVIDGYSTNPPF--NEWKKFTRG--WVF
 SVDVTVTARAGAPVFLPGPAAR--PDVRAVRCGWSVLAGA¹SPVPVEP
 --VELSEGAWAMIDGRDVLTP--NT¹TRVTKA--WTF
 SWRRVS¹VGEDVLLPAPGPTGR--GPTQKLWAVEPLDG¹CPLHPSW
 --LDLLEGQSV¹VAIDPPPLT--NG¹ITYTET--PTW
 NAEGLGDGNGNEHVDLVNRTMKNVLMKHGHRNFMDASILYKSVHGITHL

P N1 P N2 # 1 # # 2
 REAPP^{HL}HLV^{SV}SEGAN^{FL}FLDARGDGA^VLAGI^{WT}FLPVRG^{CD}DAVS^{VT}TV^{CF}
 REAPP^{HL}HLV^{SV}SEGAN^{FL}FLDARGDGA^VLAGI^{WT}FLPVRG^{CD}DAVS^{VT}TV^{CF}
 REAPP^{HL}HLV^{SV}SEGAN^{FL}FLDARGDGA^VVAGI^{WT}FLPVRG^{CD}DAVA^{VT}VM^{VC}
 LQT^{PP}SYCK^{GD}GIS^{IS}NI^{CI}ERN-----I^{ED}E^{DI}F^{LN}K^{RC}T
 VCL^{DR}DC^{FT}D^{VA}LD^{AA}CL^{RT}AR-----V^{AP}LA^{TA}E^{AL}ER^P
 LET^{PP}G⁻CAG^DIS^VKK^VVS^{HS}-----L^{ED}NI^II^GK^{HN}
 VSL^{MP}K^{QV}PET^{VD}AA^{CR}AP^{VP}-----L^{AM}AY^{AP}PA^{PS}AT^{GL}RT^D
 ISS-----I^{CD}NTSMGY^{IL}DL^{DR}-----A^{TG}F^{QD}LL^{GS}CS^V
 A^{HR}OP^{TE}VS^{VA}EN^{OL}L^{LK}VP^{HO}EN^{HT}Y^{TER}WS^{FI}PA^{AP}CK^LTP^{PS}I

ETATCHPDLVLGRGAVPEAPEMGIGDYLPPPEVRLRREPIVTPERWSPHL
ETATCHPDLVLGRGAVPEAPEMGIGDYLPPPEVRLRREPIVTPERWSPHL
ETATCHPDLVLGRGAVPEAPERGIGDYLPPPEVRLQREPIVTPERWSPHL
IKTINYPELAVADFEISNNTIKKINDVYFVNDSEVF---PIITTKSG--I
STDGKEFLVADPHVSAQLGRNATGVLIAAAAEEDGGVYFLYDLRIDGAD
LLTGEHGIALAEFNVVNGSLRRTDDVYFVNGTVF---PILATRSV--L
FVWQGEAAVNVNPSLVIHGVRETDSGLTYLSVGDIKDPARQVASVVLVQV
RYGEEKILVRDRFVNVNSDGLDRIASSQFHYIPN-----VIIGTRGKEL
DVCIKHGACIHIDVNVGDCIARSEMHPLVETGVYVHAKVADHPWTVA

```

#      PN3
SV-----LRATPNDTGLYTLHDASG---PRAVFFVA
SV-----LRATPNDTGLYTLHDASG---PRAVFFVA
TV-----RRATPNDTGLYTLHDASG---PRAVFFVA
HI-----TNVTINNSGIYTLHYENNDKWSHQSKILVT
DE-----ETQLALTQLVATAGACQGAARDEEREPATG
QI-----HRATPSIAGVYTLHVSIDGMMKHSVLLLT
AP-----VPTPPPTPADYDEDDNDEGEDESLAGLVA
TI-----FNATSQIAGVYTRYTRNDSSRPVAMDVLVV
EDFANYGFDTPDVKPGVLKFEHNNAAHGVAIWNLOGTHGENMYVTFLV

```

(continued)

B) Cysteine clusters in N-terminal

Local TK⁻gE⁺PrV
 PrV:Ea strain (AF171937)
 PrV:Rice strain (P08354)
 CHV (AAB67060)
 BHV-1:ST strain (Q08101)
 EHV-1 (M36299)
 HSV-1:strain 17 (P04488)
 GaHV-2 (CAA486)
 SVV:DHV strain (Q04548)

Local TK⁻gE⁺PrV
 PrV:Ea strain (AF171937)
 PrV:Rice strain (P08354)
 CHV (AAB67060)
 BHV-1:ST strain (Q08101)
 EHV-1 (M36299)
 HSV-1:strain 17 (P04488)
 GaHV-2 (CAA486)
 SVV:DHV strain (Q04548)

Local TK⁻gE⁺PrV
 PrV:Ea strain (AF171937)
 PrV:Rice strain (P08354)
 CHV (AAB67060)
 BHV-1:ST strain (Q08101)
 EHV-1 (M36299)
 HSV-1:strain 17 (P04488)
 GaHV-2 (CAA486)
 SVV:DHV strain (Q04548)

Local TK⁻gE⁺PrV
 PrV:Ea strain (AF171937)
 PrV:Rice strain (P08354)
 CHV (AAB67060)
 BHV-1:ST strain (Q08101)
 EHV-1 (M36299)
 HSV-1:strain 17 (P04488)
 GaHV-2 (CAA486)
 SVV:DHV strain (Q04548)

Local TK⁻gE⁺PrV
 PrV:Ea strain (AF171937)
 PrV:Rice strain (P08354)
 CHV (AAB67060)
 BHV-1:ST strain (Q08101)
 EHV-1 (M36299)
 HSV-1:strain 17 (P04488)
 GaHV-2 (CAA486)
 SVV:DHV strain (Q04548)

```

#
VGD-----RPPAPADPVGPARHEPRFHALGFHSQSFSPGDTF
VGD-----RPPAPADPVGPARHEPRFHALGFHSQSFSPGDTF
VGD-----RPPAPLAPVGPARHEPRFHALGFHSQSFSPGDTF
IKK-----KETVITKPKVYIKKHGGFFHVKNYHSHVVFVNDSEF
PTP-----GPPPHRTTTRAPPRRHGARFRVLPYHSHVYTPGDSF
VKKPPKQPQPRLRVKT PPPVTVPQVPVKTHT-DFVVHGYHSRVYADGSEF
SGTFRLLPPPP----APPSWPSAPEVSHVRGVTVRMETPEAILFSPGETF
WVSVHG-----QAPDRMTNIYITPPSTRVDLTFKNFNAFLNAGDTPV
KLNN-----SIENHIDLPAVTPKPKGAEFHTWHYHSHVFSVGETF

          PN4          5          6
DLMPRVSDMGDSRENFATLTDWYYARAPP-RCLLYYVYEP-CIYHPRAPE
DLMPRVSDMGDSRENFATLTDWYYARAPP-RCLLYYVYEP-CIYHPRAPE
DLMPRVSDMGDSRENFATLTDWYYARAPP-RCLLYYVYEP-CIYHPRAPE
KIELNLESEIYDS--EFSASIDWYMKTS--ECSVFHIYETCIFIHPHANS
LLSVRLQSEFFDEAP-FSASIDWYFLRTAG-DCALIRIYETCIFIHPEAPA
ELSVNLESHIVEP--SFSAEIQWYMNNTSSSSCDLFRVFETCIFIHPTAMA
STNVSIIHAIAHDD-QTYSDVWVLRFDVPT-SCAEMRIYESCLYHPQLPE
DASVTAHLTNWAG--AFSLEFKLLFLPYNS-KCLYLTVEYPCIFIHPSESE
SLPMHLQYKIHD--PFDLLEWLYVPINE-TCQPMRLYSACVYHETVPS

          :          *          :...*:
7          8          #          9          10
C-LRPVDPACSFSTSPARARLVARRAYASCSPLLGDRWLTACPFDAFG---
C-LRPVDPACSFSTSPARARLVARRAYASCSPLLGDRWLTACPFDAFG---
C-LRPVDPACSFSTSPARALVARRAYASCSPLLGDRWLTACPFDAFG---
C-LNPINPLCSFTSPLRATSLINRFYFRCKPE-GKNWTTDCINTYSINAD
C-LHPADAQCSFASPYRSETVYSRLYECCRPDPACGRWPHECEGAAYAAPV
C-LHPEQHTCSFTSPIRATKILHRVYGNCSDH-GNSWSPSRCHSTLLGNRL
C-LSPADAECAAST--WTSRLAVRSYAGCSR---TNPPPPCSAEAHMEPV
COMVPEHAECRFAANMDVMQLASARTDGCRRVDVRGCTFQTSVDESVOGKL
C-LSPENPECTFASPHIARRVANTVYQNCHE---VNYTADCLAVSHVEPG
* : * . * : :          :          *

          PN5
-----EEVHTNATADESGLYVLVMTNNGHVATWDYTLVATAAEYV
-----EEVHTNATADESGLYVLVMTNNGHVATWDYTLVATAAEYV
-----EEVHTNATADESGLYVLVMTNNGHVATWDYTLVATAAEYV
KHIKQHSNNVDLIFLNTPTNASGLYVFILKYNHPEAWTYTLVSTVKNFM
AHLRPANNSVDLVFDDAPAAASGLYVFLQYNGHVEAWDYSLVVTSDRLV
YFIQPAQNVRVDLLFKDTPASATGLYVFLVLYNGHPEAWTYTLSTANHFV
PGLAWQAASVNLEFRDASPOHSGLYLCVVYVNDHIHAWGHITISTAAQYR
AFLESVDP--SFRLANAQPTDAGLYVIVGLYNGRPLAWTYVYLSTLETIL
SGLEIQNGGSALLFVNAAESMSGLYVFIHFNGHVETVAYTVVSTIENFV
: :          :***: :          * : : : : *

#
TVIKELTAPARAPGTPWGPGGGDDAIYVDGVTTTAPPARPNWNPYGRITTPG
TVIKELTAPARAPGTPWGPGGGDDAIYVDGVTTTAPPARPNWNPYGRITTPG
TVIKELTAPARAPGTPWGPGGGDDAIYVDGVTTTAPPARPNWNPYGRITTPG
NVIKDMTRPLLSNNKMKKPEHSTQPPTITNITPGFKSKNWVDKY-----
RAVTDHTRPEAAAAAPEPGPPLTSEPAAPTGPAP-----
NVLTDVTRPRLGEHFYTDLGHKIITPHPS-VATTEELGANTRHY-----
NAVVEQPLPQRGADLAEPHVPVHGAAPPHTHGA-----
NVFEDVHKPGFGYNAVSADTPGNTAAPTHTSAFKEGP-----T
NAIEEHGFPPEIHNVPSPSSPNVTANNDVISETNTFP-----F
.. :          *

```

(continued)

C) Transmembrane sequences

Local TK ⁻ gE ⁺ PrV	RLFVLALGSEFVMTVCVVGAVWLCVLCSSRRRAASRPFRVPTRVRTHMLSP
PrV:Ea strain (AF171937)	RLFVLALGSEFVMTVCVVGAVWLCVLCSSRRRAASRPFRVPTRARHMLSP
PrV:Rice strain (P08354)	RLFVLALGSEFVMTVCVVGAVWLCVLCSSRRRAASRPFRVPTRAGTRMLSP
CHV (AAB67060)	IISVAVVSCITIVILIVVITFCVHCIG--LNRKPYEIIIN-----PFNT
BHV-1:ST strain (Q08101)	-WLVLVLGALGLAGLVGIAALAVRVCARRASQKRTYDILN-----PFGP
EHV-1 (M36299)	LAFLVICTCAALLVALVWGCILYIR--SNRKPYEVLN-----PFET
HSV-1:strain 17 (P04488)	LRLGAVMGAALLLSALGLSVWACMTCWRRRAWRAVKSRA--GKGP
GaHV-2 (CAA486)	VIYSLLVSSMAAGVILVLLLALLIVGLYKRHARHRTNGY---FQAYP
SVV:DHV strain (Q04548)	KTYAGITGGFAVLALVCLALALVCTKRKFCHRSYWSDKAAYG---QST
	:
Local TK ⁻ gE ⁺ PrV	YTSPLTHEDYYDGD--DDDEEAGVRRRPPASPGGDSGYEGPYTSLDPEDE
PrV:Ea strain (AF171937)	YTSPLTHEDYYDGD--DDDEEAGVIRRRPPASPGGDSGYEGSYASLDPEDE
PrV:Rice strain (P08354)	YTSPLTHEDYYDGD--DDDEEAGDARRRPSSPGGDSGYEGPYVSLDAEDE
CHV (AAB67060)	YKSIPTNEKN--IL-HFAEVTESDYSSDESFDSDS--EELNQRG-----
BHV-1:ST strain (Q08101)	YTSPLTNEPLDVVPVSDDEFSLDEDSFADDDSDDDGPASNPPADAYDL
EHV-1 (M36299)	YTSVPSNDPSDEVL-VFERLAS---DSDDSFSDSDSELEYPPP-PKPA
HSV-1:strain 17 (P04488)	YIRVADSELYADWS--SDSEGERDQVPWLAPPERPDSPTNGSGFEILS
GaHV-2 (CAA486)	YSSLPSNDECVFGD---GDFNSPLSNTCEGLSRGLAGNSK-----
SVV:DHV strain (Q04548)	YAGVPVDDFEDDTEVEVDEG-----
	* : :
Local TK ⁻ gE ⁺ PrV	SSDEDDG---LYVRPEEAPRSGFDVWFRDPEKPE-----VTTGPNY
PrV:Ea strain (AF171937)	SSDEDDG---LYVRPEEAPRSGFDVWFRDPEKPE-----VTNGPNY
PrV:Rice strain (P08354)	SSDEDDG---LYVRPEEAPRSGFDVWFRDPEKPE-----VTNGPNY
CHV (AAB67060)	-----ET-----IQQKKE--QSGYTIWFNEDLEES-----VSKKLNQPNY
BHV-1:ST strain (Q08101)	GAPEPTSGFARAPANGTIRSSRSGFKVWFRDPPEDD--AAPARAPAPDY
EHV-1 (M36299)	QLPPYQF---VDGGDAPSGRSGFKVWFRDTPEASPVPLHKPTLQGPDY
HSV-1:strain 17 (P04488)	TAPSVYP-----RSDGHQSRRLTTFGSGRPDRR-----Y
GaHV-2 (CAA486)	-----SRTPKQKGSRYHAWFADG-----GPAA
SVV:DHV strain (Q04548)	-----AECGGSGYTVYIDKRTR-----
	:
Local TK ⁻ gE ⁺ PrV	VTANRLILMSRPA
PrV:Ea strain (AF171937)	VTANRLILMSRPA
PrV:Rice strain (P08354)	VTASRLLNARPA
CHV (AAB67060)	KIINSLKSIQNE
BHV-1:ST strain (Q08101)	VVAARLKSILR-
EHV-1 (M36299)	RVASKLKSILK-
HSV-1:strain 17 (P04488)	QASDSSVFW---
GaHV-2 (CAA486)	IRRREV-----
SVV:DHV strain (Q04548)	-----

Figure 3: Amino acid sequence alignments of the gE gene of TK⁻gE⁺ PrV and homologous proteins. The bracketed accession number indicates the retrieved location in GenBank. Perfectly conserved single amino acids are indicated with an (*); conserved strong groups as (:); conserved weak groups as (.); and none consensus as (-). A) represents the signal sequences (grey shaded) that were defined using the SignalP server; black shaded cysteine clusters particularly in PrV strains (1-4). B) particularly indicates the well conserved cysteine clusters at the C-terminal (black shaded: 5-10). C) Amino acids variations in PrV strains are indicated as #. Potential N-glycosylation sites in PrV gE sequences are numbered PN1-N5, while in the other homologous, highlighted in blue shades.

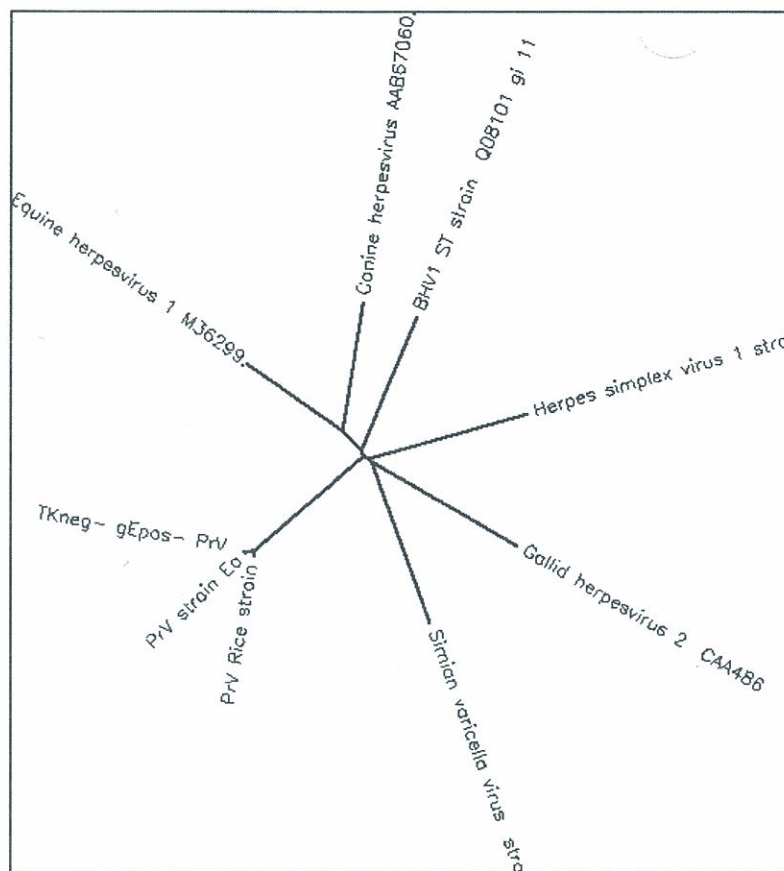


Figure 4: Phylogenetic relationships among alphaherpesvirus gE homologous proteins

show highest degree of alignment to these gE homologous protein. Indeed, by insertion of minor gaps in the multiple sequence alignment, more cystein residues could be clustered as PrV gE counterparts, indicating the conservation of some important sequences for the structure of these glycoproteins (Figure 3A).

The asparagines (N) which may be the sites of N-linked glycosylation in gE protein of TK⁻gE⁺ PrV are located at positions aa 88, 94, 180, 259 and 344 (Figure 2). The five predicted sites are closely conserved among the PrV gE sequences (Figure 3). There are a considerable number of N-linked glycosylation sites in CHV (9 sites), GaHV-2 (5 sites), EHV-1 (4 sites), SVV (4 sites), HSV-1 (3 sites) and the least in BHV-1 gE (2 sites). A single site in both GaHV-2 and CHV (PN1 and PN3 in Figure 3) aligned with that of gE PrV, but, the N-linked glycosylation consensus tripeptide differed. Therefore, the gE of PrV strains have no counterparts in terms of conservation of N-linked glycosylation sites.

DISCUSSION

This study contributes to the first documented gE sequence analysis of a Malaysian derived PrV. The ORF which codes a polypeptide of 578 amino acids presents typical characteristics compatible with the structure of a

viral glycoprotein: signal peptide, putative glycosylation sites and a long C-terminal transmembrane alpha-helix. The ORF of PrV strains revealed a striking collinearity and a highest degree of sequence conservation when compared to their homologues in other alpha herpesviruses. Besides the six revealed amino acids substitution, otherwise, the gE proteins encoded by the Ea PrV strain, turned out to be the closest relatives of TK⁻gE⁺ PrV. This may point to a close cognate relationship. Meanwhile, codon usage analysis indicated a bias to GC-rich codon, which is in parallel to previous documentation (Petrovskis *et al.*, 1986). Although codon usage is least well understood in higher organisms, significant correlation to gene expression levels, tissue-specific patterns of expression, the degree of evolutionary conservation of proteins, and the overall or regional nucleotide composition of the genome had been evidenced in herpesviruses (Cristillo *et al.*, 2001; Porter, 1995). In order for any secondary structural element to form in particularly, the gE gene, it is necessary to have most amino acids within its sequence to have a high propensity for that structure. The significant conservation of GC-rich codons observed among the analysed PrV gE sequences could address this notion. Identical aforementioned GC-rich amino acids would therefore cluster over these length ranges as this would favour a

TK ⁻ gE ⁺ PrV	H2N-117-C-8-C-4-C-9-C-132-C- 8-C- 8-C-7-C-18-C-11-C-109-C-8CVLC at 451-124-COOH
PrV Ea	H2N-117-C-8-C-4-C-9-C-132-C- 8-C- 8-C-7-C-18-C-11-C-109-C-8CVLC at 451-124-COOH
PrV Rice	H2N-116-C-8-C-4-C-9-C-132-C- 8-C- 8-C-7-C-18-C-11-C-109-C-8CVLC at 450-124-COOH
CHV	H2N-12-C-50-C-9-C-5-C-9-C-143-C- 8-C- 8-C-7-C-18-C-10-C-111-C-12-C-3-C-103-COOH
SVV	H2N-14-C-176-C-10-C-5-C-9-C-150-C- 8-C- 8-C-7-C-18-C-8-C-113-C-6-C-44-C-14-COOH
BHV-1	H2N-75-C-8-C-5-C-9-C-168-C- 8-C- 8-C-7-C-18-C-11-C-119-C-128-COOH
EHV-1	H2N-9-C-54-C-9-C-5-C-9-C-155-C- 8-C- 8-C-7-C-18-C-10-C-110 CTC at 414-11-C-122-COOH
GaHV-2	H2N-3-C-54-C-8-C-5-C-9-C-147-C- 8-C- 8-C-8-C-18-C- 6-C-156-C-14-C-41-COOH
HSV-1	H2N-13-C-3-C-44-C-24-C-182-C- 8-C- 8-C-7-C-16-C- 8-C-35-C-80 CMTC at 440-107-COOH

Figure 5: Spacing of cystein residues of PrV gE and homologous proteins. The number on the right of each cystein (C) residue indicated the position in the amino acid sequence. The N-terminus and C-terminus of the sequence are represented by H₂N and COOH, respectively. Red font: highly conserved region; Bold: moderately conserved

sequence with a high preference for forming one particular secondary structure. For instance, alanine has a high preference for the α -helix. Hence evolution will select sequences where alanines are clustered in order to favour α -helix formation as presented in the study. If amino acids were randomly distributed, the probability that a stretch of amino acids would contain a high preference for a secondary structural element would be decreased. The characteristic nature of PrV gE and homologous proteins as a type-1 membrane protein clearly evidenced the regional gene conservation where amino acids of similar hydrophobicity clustered in order to produce a hydrophobic membrane spanning sequence or water-exposed polar loop.

Based on the cystein similarities, all the 10 clusters in the gE sequence of TK⁻gE⁺ PrV, Ea strain and Rice strain were strictly conserved. These clusters are likely to form intramolecular disulfide bridges which are important for the folding and function of the gE protein (Fariselli *et al.*, 1999). Despite the low overall level of amino acid sequence identity among the gE proteins of the diverse animal species, which is on the order of 23 to 31%, the cystein clusters were relatively well conserved but the best homology resides in the third C- terminal part of the protein where 6 cystein residues could be aligned to the homologues of BHV-1, CHV, EHV-1, HSV-1, SVV with 5 conservations in GaHV-2. The strong conservation of cystein residues amongst all alphaherpesviruse gE sequences investigated implied some degree of conservation of the secondary and tertiary structure of the proteins.

From the sequence analysis, TK⁻gE⁺ PrV gE protein sequence contains five potential N-glycosylation sites,

which are well conserved both in position and number among the different PrV strains. Surprisingly, the sites are not conserved in other homologous gE protein sequence. Although N-glycosylation consensus sequence, NXT/S, also known as the sequon, are abundant in proteins, but only two-thirds are glycosylated, due to the fact that the folding of the protein has enormous implication in the N-glycosylation regulation (Pless and Lennarz, 1977). Any mutation in the tripeptide consensus sequence would lead to a nonglycosylated form of protein. For example, attenuated PrV strain Bartha (PrV-Ba) expresses a nonglycosylated gM (Dijkstra *et al.*, 1997) due to a point mutation in the DNA sequence specifying the sole conserved consensus motif for N-glycosylation. Little is known, however, about the influence of other residues at this position, nor of those flanking the sequon, on the efficiency of N-glycosylation (Shakin-Eshelman, 1996). The lack of glycosylation of some sequons may be the result, at least in part, of the presence or absence of specific residues at or near the sequon. For instance, a proline (Pro) at the X position was reported to be prohibitive for glycosylation (Gavel and von Heijne, 1990).

CONCLUSIONS

In conclusion, the study revealed a significant homology of gE of TK⁻gE⁺ PrV to PrV Ea and PrV Rice strains. Comparison of amino acid sequences with gE counterparts indicated a greater diversity of sequence in the N- terminal region of the protein and highlighted several features of the gE protein conserved throughout the herpesvirus family.

REFERENCES

- Card, J.P., Whealy, M.E., Robbins, A.K. and Enquist, L.W. (1992). Pseudorabies virus envelope glycoprotein gI influences both neurotropism and virulence during infection of the rat visual system. *J. Virol.* **66**: 3032-3041.
- Cristillo, A.D., Mortimer, J.R., Barrette, I.H., Lillicrap, T.P. and Forsdyke, D.R. (2001). Double-stranded RNA as a not-self alarm signal: to evade, most viruses purine-load their RNAs, but some (HTLV-1, Epstein-Barr) pyrimidine-load. *J. Theoretical. Biol.* **208**: 475-489.
- Dijkstra, J.M., Mettenleiter, T.C. and Klupp, B.G. (1997). Intracellular processing of pseudorabies virus glycoprotein M (gM): gM of strain Bartha lacks N-glycosylation. *Virology* **237**: 113-122.
- Enquist, L.W., Husak, P.J., Banfield, B.W. and Smith, G.A. (1999). Infection and spread of alphaherpesviruses in the nervous system. *Adv. Virus. Res.* **51**: 237-347.
- Fariselli, P., Riccobelli, P. and Casadio, R. (1999). Role of evolutionary information in predicting the disulfide-bonding state of cysteine in proteins. *Proteins* **36**: 340-346.
- Gavel, Y. and von Heijne, G. (1990). Cleavage-site motifs in mitochondrial targeting peptides. *Prot. Eng.* **4**: 33-37.
- Jacobs, L., Moonen-Leusen, B.M.W.M., Bianchi, A.T.J. and Kimman, T.G. (1993a). Glycoprotein I of pseudorabies virus: epitope-specific antibody response in mice and pigs. In: Abst. Int. Symp. Aujeszky's Disease, Budapest, Hungary, pp. 11.
- Jacobs, L., Rziha, H.J., Kimman, T.G., Gielkens, A.L.J. and Van Oirschot, J.T. (1993b). Deleting valine-125 and cysteine-126 in glycoprotein gI of pseudorabies virus strain NIA-3 decreases plaque size and reduces virulence in mice. *Arch. Virol.* **131**: 251-264.
- Kimman, T.G., De Wind, N., Oei-Lie, N., Pol, J.M.A., Berns, A.J.M. and Gielkens, A.L.J. (1992). Contribution of single genes within the unique short of Aujeszky's disease virus (suid herpesvirus type 1) to virulence, pathogenesis and immunogenicity. *J. Gen. Virol.* **73**: 243-251.
- Petrovskis, E.A., Timmins, J.G. and Post, L.E. (1986). Use of lgt11 to isolate genes for two pseudorabies virus glycoproteins with homology to herpes simplex virus and varicella-zoster virus glycoproteins. *J. Virol.* **60**: 185-193.
- Pless, D.D. and Lennarz, W.J. (1977). Enzymatic conversion of proteins to glycoproteins. *Proc. Natl. Acad. Sci. USA.* **74**: 134-138.
- Porter, T.D. (1995). Correlation between codon usage, regional genomic nucleotide composition, and amino acid composition in the cytochrome P-450 gene superfamily. *Biochim. Biophys. Acta.* **1261**: 394-400.
- Schoenbaum, M.A., Zimmerman, J.J., Beran, G.W. and Murphy, D.P. (1990). Survival of pseudorabies virus in aerosol. *Am. J. Vet. Res.* **51**: 331-333.
- Shakin-Eshelman, S.H., Spitalnik, S.L. and Kasturi, L. (1996). The amino acid at the X position of an Asn-X-Ser sequon is an important determinant of N-linked core-glycosylation efficiency. *J. Biol. Chem.* **271**: 6363-6366.
- Tirabassi, R. S., and Enquist, L.W. (1999). Mutation of the YXXL endocytosis motif in the cytoplasmic tail of pseudorabies virus gE. *J. Virol.* **73**: 2717-2728.
- Tirabassi, R.S., Townley, R.A., Eldridge, M.G. and Enquist, L.W. (1997). Characterisation of pseudorabies virus mutants expressing carboxy-terminal truncations of gE: evidence for envelope incorporation, virulence and neurotropism domains. *J. Virol.* **71**: 6455-6464.
- Zeenathul, N.A. (1999). Characterization of pseudorabies virus and its mutants. MSc thesis. Universiti Putra Malaysia, Malaysia.
- Zeenathul, N.A. (2004). Development of gene deleted recombinant pseudorabies virus. PhD thesis. Universiti Putra Malaysia, Malaysia.